

### AMENDMENTS TO THE SPECIFICATION

Please revise the substitute specification provided on December 21, 2006 to include the following amendments.

Please amend Table 69 of paragraph [00323] at pages 268-269 as shown below. Please insert a number “4” after the heading to represent the new footnote 4 being added to the specification, as shown below:

<b>Table 69. Gene expression signatures associated with recurrent prostate cancer.</b> (SEQ ID NOS 2240-2265, respectively, in order of appearance) <sup>d</sup>			
<b>Table 69a: Signature 1</b>			
<b>LocusLink Name</b>	<b>Gene Name</b>	<b>GenBank <u>GENBANK®</u> ID</b>	<b>UniGene ID</b>
MGC5466	Hypothetical protein MGC5466	U90904	Hs.83724
Wnt5A	proto-oncogene Wnt5A	L20861	Hs.152213
KIAA0476	KIAA0476 protein	AB007945	Hs.6684
ITPR1	inositol 1,4,5-trisphosphate receptor, type 1	D26070	Hs.198443
TCF2	transcription factor 2, hepatic	X58840	Hs.169853
<b>Table 69b: Signature 2</b>			
<b>Gene</b>	<b>Gene Name</b>	<b>GenBank <u>GENBANK®</u> ID</b>	<b>UniGene ID</b>
MGC5466	Hypothetical protein MGC5466	U90904	Hs.83724
CHAF1A	Chromatin assembly factor 1, subunit A	U20979	Hs.79018
CDS2	CDP-diacylglycerol synthase 2	Y16521	Hs.24812
IER3	Immediate early response 3	S81914	Hs.76090
<b>Table 69c: Signature 3</b>			

<b>LocusLink Name</b>	<b>Gene Name</b>	<b>GenBank GENBANK® ID</b>	<b>UniGene ID</b>
PPFIA3	Protein tyrosine phosphatase, receptor type, f polypeptide	AB014554	Hs.109299
COPEB	Core promoter element binding protein	AF001461	Hs.285313
FOS	V-fos oncogene homolog	V01512	Hs.25647
JUNB	Jun B proto-oncogene	X51345	Hs.400124
ZFP36	zinc finger protein 36, C3H type	M92843	Hs.343586

Please also amend the footnote section at the bottom of page 269, to insert a new footnote 4, as follows:

<sup>4</sup> As explained in footnote 1, correspondence between the rows in the Tables and the SEQ ID NOS listed in the Table heading of each Table is established as follows: each set of related, consecutive sequences (e.g., nucleotide sequences encoding a protein or homologous nucleotide sequences) is assigned to a single row in the Table, and the next such set is assigned to the following row in the Table going down the rows in the Table. Where only the nucleotide sequence was available for a gene, that gene includes only one SEQ ID NO. For example, in Table 69, SEQ ID NO 2240 represents the nucleotide sequence for the first horizontal row of Table 69a showing “MGC5466,” and there is no amino acid sequence for this gene. SEQ ID NOS 2241 and 2242 represent the nucleotide and amino acid sequences, respectively of the horizontal row corresponding to “Wnt5A.” SEQ ID NOS 2243 and 2244 represent the nucleotide and amino acid sequences, respectively of the horizontal row corresponding to “KIAA0476.” SEQ ID NOS 2245 and 2246 represent the nucleotide and amino acid sequences, respectively of the horizontal row corresponding to “ITPR1.” SEQ ID NOS 2247 and 2248 represent the nucleotide and amino acid sequences, respectively of the horizontal row corresponding to “TCF2.” SEQ ID NO 2249 represents the nucleotide sequence of “MGC5466” in Table 69b, and there is no amino acid sequence for this gene. SEQ ID NOS 2250 and 2251 represent the nucleotide and amino acid sequences, respectively of the horizontal row corresponding to

“CHAF1A,” and so forth continuing with SEQ ID NOS 2252-2253 corresponding to “CDS2,” SEQ ID NOS 2254-2255 corresponding to IER3, SEQ ID NOS 2256-2257 corresponding to “PPF1A3” in Table 69c, SEQ ID NOS 2258-2259 corresponding to “COPEB,” SEQ ID NOS 2260-2261 corresponding to “FOS,” SEQ ID NOS 2262-2633 corresponding to “JUNB,” and SEQ ID NOS 2264-2265 corresponding to “ZFP36.”

Please amend the following row of Table 8 on p. 106 as follows:

38288_at	N/A [ <del>GenBank</del>  GENBANK® Accession  No. L42611]	KRT6E: keratin 6E
----------	---	-------------------

Please amend the following heading row of Table 29 of p. 148 as follows:

<b>Table 29. Breast Cancer Good Prognosis Minimum Segregation Set 2.</b>					
(SEQ ID NOS 586-608, respectively, in order or appearance)					
<b>r = - 0.992</b>	<b>System 1 (MCF7)</b>				
<b>Locus Link</b>	<del>GenBank</del>	<b>UniGene</b>	<b>Systematic</b>	<b>Gene</b>	<b>Gene</b>
<b>Symbol</b>	<u>GENBANK®</u>		<b>name</b>	<b>name</b>	<b>Description</b>

Please amend the following heading row of Table 41 of p. 143 as follows:

<b>Table 41. The 9-gene molecular signature associated with metastatic prostate cancer</b>			
(SEQ ID NOS 1255-1271, respectively, in order or appearance)			
<b>Gene</b>	<b>Gene name</b>	<del>GenBank</del> <u>GENBANK® ID</u>	<b>UniGene ID</b>

Please amend the following heading row of Table 43 of p. 195 as follows:

<b>Table 43. The 5-gene molecular fingerprint associated with invasive phenotype of human prostate cancer</b>	
(SEQ ID NOS 1272-1280, respectively, in order or appearance)	

Gene	Gene name	GenBank <u>GENBANK®</u> ID	UniGene ID
------	-----------	----------------------------	------------

Please amend the following heading row of Table 44 of p. 195 as follows:

<b>Table 44. The 8-gene molecular fingerprint predicting recurrent phenotype of human prostate cancer</b> (SEQ ID NOS 1281-1295, respectively, in order or appearance)			
Gene	Gene name	GenBank <u>GENBANK®</u> ID	UniGene ID

Please amend the following heading row of Table 45 of p. 197 as follows:

<b>Table 45. The 25-gene molecular signature predicting recurrent prostate cancer</b> (SEQ ID NOS 1296-1347, respectively, in order or appearance)			
Gene	Gene name	GenBank <u>GENBANK®</u> ID	UniGene ID

Please amend the following heading row of Table 46 of p. 199 as follows:

<b>Table 46. The 12-gene molecular signature predicting recurrent prostate cancer</b> (SEQ ID NOS 1348-1369, respectively, in order or appearance)			
Gene	Gene name	GenBank <u>GENBANK®</u> ID	UniGene ID

Please amend paragraph 112, pp. 31-32, as follows:

[00112] For example, consider an embodiment in which the first reference set is derived using data obtained from three separate control cell lines and six separate tumor cell lines. For each gene considered for inclusion within the first reference set, pairwise comparisons are carried out for each of the 3 x 6 or 18 pairwise combinations between control cell lines and tumor cell lines. A candidate gene will be included in the first reference set if each of the 18 pairwise

comparisons reveals the gene to be consistently differentially expressed (*i.e.*, gene expression always is higher in the control cell line or always higher in the tumor cell line for each of the 18 pairwise comparisons). As one of ordinary skill readily will appreciate, it may sometimes be necessary to scale the datasets prior to carrying out the pairwise comparisons. Such scaling may be routinely implemented in the analysis software provided by commercial suppliers of expression arrays or array readers (such as, *e.g.*, ~~Affymetrix~~AFFYMETRIX®, Santa Clara, CA). For a general discussion of data scaling for and differential gene expression analysis, *see, e.g.*, ~~Affymetrix~~AFFYMETRIX® Microarray Suite 4.0 User Guide, ~~Affymetrix~~AFFYMETRIX®, Santa Clara, CA, incorporated herein by reference.

Please amend paragraph 172, pp. 78-79, as follows:

[00172] ~~Affymetrix~~AFFYMETRIX® **arrays**. The protocol for mRNA quality control and gene expression analysis was that recommended by the array manufacturer, ~~Affymetrix~~AFFYMETRIX®, Inc. (Santa Clara, CA, *see* ~~Affymetrix~~AFFYMETRIX® website). In brief, approximately one microgram of mRNA was reverse transcribed with an oligo(dT) primer that has a T7 RNA polymerase promoter at the 5' end. Second strand synthesis was followed by cRNA production incorporating a biotinylated base. Hybridization to ~~Affymetrix~~AFFYMETRIX® Hu6800 arrays representing 7,129 transcripts or ~~Affymetrix~~AFFYMETRIX® U95Av2 array representing 12,626 transcripts overnight for 16 h was followed by washing and labeling using a fluorescently labeled antibody. The arrays were read and data processed using ~~Affymetrix~~AFFYMETRIX® equipment and software (Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. Expression monitoring by hybridization to high-density oligonucleotide arrays [*see* comments]. *Nat. Biotechnol.* 1996;14:1675-80, incorporated herein by reference). Detailed protocols for data analysis and documentation of the sensitivity,

reproducibility and other aspects of the quantitative microarray analysis using Affymetrix AFFYMETRIX® technology have been reported (Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. Expression monitoring by hybridization to high-density oligonucleotide arrays [*see comments*]. Nat. Biotechnol. 1996;14:1675-80, incorporated herein by reference).

Please amend paragraph 173, pp. 79-80, as follows:

[00173] To determine the quantitative difference in the mRNA abundance levels between two samples, in each individual sample for each gene the average expression differences were calculated from intensity measurements of perfect match (PM) probes minus corresponding control probes representing a single nucleotide mismatch (MM) oligonucleotides for each gene-specific set of 20 PM/MM pairs of oligonucleotides, after discarding the maximum, the minimum, and any outliers beyond 3 standard deviations (SD) from the average. The averages of pairwise comparisons for each individual gene were made between the samples, and the corresponding expression difference calls (*see below*) were made with Affymetrix AFFYMETRIX® software. Microsoft Access was used for other aspects of data management and storage. For each gene, a matrix-based decision concerning the difference in the mRNA abundance level between two samples was made by the software and reported as a "Difference call" (No change (NC), Increase (I), Decrease (D), Marginal increase (MI), and Marginal decrease (MD)) and the corresponding fold change ratio was calculated. 40-50% of the surveyed genes were called present by the Affymetrix AFFYMETRIX® software in these experiments. The concordance analysis of differential gene expression across the data set was performed using Microsoft Access and Affymetrix AFFYMETRIX® MICRODB software. For experiments involving study of prostate cancer, three of the normal prostate epithelial (NPE)

microarrays are used as controls, and referred to as the NPE expression profile. Thus, when a gene is required to show a 2-fold or greater change relative to NPE, this must occur in all three microarrays, for either positive or negative changes. These stringent criteria exclude genes for which one of the three microarrays is in error. The strategy in this study is based on the idea that expression differences will not be called by chance in the same direction in multiple arrays (see below for statistical justification). Each gene in the final list of the 214 differentially expressed genes was required to be called exclusively as either concordantly up- or down-regulated in 30 separate comparisons (5 prostate cancer cell lines x 2 experimental serum conditions x 3 NPE controls) or 15 separate comparisons (5 prostate cancer cell lines x 1 experimental serum condition x 3 NPE controls).

Please amend paragraph 174, as follows:

[00174] **Statistical analysis and quality performance criteria.** We used a stringent analytical approach to test the hypothesis that there are common genes with altered mRNA abundance levels which appear to be significantly associated with the studied phenotypes. The ~~Affymetrix~~ **AFFYMETRIX®** MICRODB and ~~Affymetrix~~ **AFFYMETRIX®** DMT software was used to identify in any given comparison of two chips only genes that are determined to be expressed at statistically significantly different ( $p < 0.05$ ) levels. These transcripts are called as differentially expressed. To be included in our final differentially regulated gene class the given transcript was required to be determined as differentially regulated in the same direction (up or down) at the statistically significant levels ( $p < 0.05$ ) e.g., in 30 independent comparisons (5 experimental cell lines X 2 experimental conditions X 3 control cell lines). To be recognized as differentially regulated in the orthotopic tumors any given gene of the PC3/LNCap consensus class was required to be determined differentially regulated in the same direction at the statistically significant level ( $p < 0.05$ ) in 18 additional independent comparisons (6 orthotopic

tumors X 3 controls). Despite that identified set of 214 genes is differentially expressed in described experimental systems with the extremely high level of confidence, we carried out Q-PCR confirmation analysis for a sub-set of identified genes and confirmed their differential expression in all instances using an additional independent normal human prostate epithelial cell line as a control.

Please amend paragraph 175 as follows:

[00175] Quality performance criteria adopted for the ~~Affymetrix~~AFFYMETRIX® GENECHIP system and applied in this study. 40-50% of the surveyed genes were called present by the Affymetrix software in these experiments. This is at the high end of the required standard adopted in many peer-reviewed publications using the same experimental system. Transcripts that are called present by the ~~Affymetrix~~AFFYMETRIX® software in any given experiment were determined to have the signal intensities higher in the perfect match probe sets compared to single-nucleotide mismatch probe sets and background at the statistically significant level. This analysis was performed for each individual transcript using unique set of 20 perfect matches versus 20 single nucleotide mismatch probes. In our final list of 214 genes all transcripts were called present in at least one experimental setting. The inclusion error associated with two mRNA samples from identical cell lines was 2.7% for a difference called by the ~~Affymetrix~~AFFYMETRIX® software. Thus, two independently obtained mRNA from the same cell lines will have 2.7% false positives. When a third independently derived epithelial cell line was included, only 4 genes (0.06%) out of 7,129 were called differentially expressed. The expression profiles of the normal prostate epithelial cell lines used in our experiments were determined to be indistinguishable. Therefore, controls are not likely source of errors in gene expression analysis performed in this study. This is particularly important, since the strategy adopted in this study is based on the idea that expression differences will not be called



statistically significant by chance in the same direction in multiple arrays and during multiple independent comparisons of different phenotypes and variable experimental conditions. To impose additional stringent restrictions on possibility of a gene to be detected as concordantly differentially regulated by chance, we apply the use of multiple experimental models and vastly variable experimental settings such as in vitro and in vivo growth and varying growth conditions. Similar strategy for identification of consistent gene expression changes based on a concordant behavior of the differentially regulated genes using ~~Affymetrix~~ AFFYMETRIX® GENECHIP system and software was applied and validated in several peer-reviewed published papers (see for example, Lee CK, Klopp, RG, Weindruch, R, Prolla, TA. Gene expression profile of aging and its retardation by caloric restriction. Science 1999; 285: 1390-1393; Ishida, S, Huang, E, Zuzan, H, Spang, R, Leone, G, West, M, Nevins, JR. Role for E2F in control of both DNA replication and mitotic function as revealed from DNA microarray analysis. Mol Cell Biol 2001; 21: 4684-4699, incorporated herein by reference). We applied more stringent criteria in our study requiring a concordance in at least 30 of 30 experiments compared to 6 of 6 comparisons in (Lee CK, Klopp, RG, Weindruch, R, Prolla, TA. Gene expression profile of aging and its retardation by caloric restriction. Science 1999; 285: 1390-1393, incorporated herein by reference); and 4 of 6 comparisons in (Ishida, S, Huang, E, Zuzan, H, Spang, R, Leone, G, West, M, Nevins, JR. Role for E2F in control of both DNA replication and mitotic function as revealed from DNA microarray analysis. Mol Cell Biol 2001; 21: 4684-4699, incorporated herein by reference). Ishida, *et al.* (Ishida, S, Huang, E, Zuzan, H, Spang, R, Leone, G, West, M, Nevins, JR. Role for E2F in control of both DNA replication and mitotic function as revealed from DNA microarray analysis. Mol Cell Biol 2001; 21: 4684-4699, incorporated herein by reference) provided a formal statistical justification that four or more concordant calls out of six comparisons cannot be explained by chance, with the probability in the range of  $1$  in  $10^{-4}$ .

Please amend paragraph 178 as follows:

[00178] The expression data were obtained using an ~~Affymetrix~~AFFYMETRIX® Human Genome-U95Av2 (“HG-U95Av2”) expression array chip (~~Affymetrix~~AFFYMETRIX®, Santa Clara, CA). The HG-U95Av2 Array represents approximately 10,000 full-length genes. Data were obtained from the HG-U95Av2 according to the manufacturer’s suggested protocols, as outlined in the Materials & Methods Section above

Please amend paragraph 179 as follows:

[00179] The original data set thus comprised a total of eight separate sets of gene expression data, five from the set of tumor cell lines and three from the set of epithelial cell lines. Fifteen separate pairwise comparisons were carried out to identify a first reference set of genes that were differentially expressed in the tumor cell lines and the epithelial cell lines. Differential expression was determined using ~~Affymetrix~~AFFYMETRIX®’s MICROARRAY SUITE software (versions 4.0 and 5.0). To be included in the first reference set, a candidate gene needed to meet two criteria: 1) the candidate gene was shown to be differentially expressed in each of the 15 pairwise comparisons; and 2) the direction of the differential (*i.e.* greater expression in the tumor cell lines *cf.* the epithelial cell lines or *vice-versa*) was consistent in each of the 15 pairwise comparisons. The first reference set comprised of 629 genes.

Please amend paragraph 181 as follows:

[00181] ~~Affymetrix~~AFFYMETRIX® MICRODB (version 3.0) and ~~Affymetrix~~AFFYMETRIX® Data Mining Tools (DMT) (version 3.0) data analysis software were used to identify genes that were differentially regulated in recurrence group compared to relapse-free group of patients at the statistically significant level ( $p < 0.05$ ; Student T-test). Candidate genes were included in the second reference set if they were identified by the DMT software as having p values of 0.05 or

less both for up-regulated and down-regulated genes. 316 genes were identified as being members of the second reference set

Please amend paragraph 200 as follows:

[00200] The methods of the invention were used along with the data reported by Singh, *et al.* (2002) to identify gene clusters associated with an invasive phenotype. Invasive phenotype was assessed by determining the presence or absence of positive surgical margins. The same first reference set described above in part A was used to generate the concordance and minimum segregation sets for invasiveness. The second reference set was obtained following the procedures described above in part B, using the supplemental data reported in Singh, *et al.* (2002) for fourteen invasive and 38 non-invasive human prostate tumors. Thus, the second reference set was obtained by using the ~~Affymetrix~~ AFFYMETRIX® MICRODB (version 3.0) and ~~Affymetrix~~ AFFYMETRIX® DATA MINING TOOLS (DMT) (version 3.0) data analysis software to identify genes that were differentially regulated in invasion group compared to non-invasive group of patients at the statistically significant level ( $p < 0.05$ ; Student T-test). Candidate genes were included in the second reference set if they were identified by the DMT software as having p values of 0.05 or less both for up-regulated and down-regulated genes. 3869 genes were identified as being members of the second reference set

Please amend paragraph 204 of p. 107 as follows:

[00204] Note that three entries in the table correspond to the same genes, i.e., 34853\_at, 209\_at, and 115\_at. They most likely represent the splice variants of the same gene (Hs.31989). According to ~~Affymetrix~~ AFFYMETRIX® annotation, the 34853\_at is an alternative splice 3 variant of the FGFR2

Please amend paragraph 206 as follows:

[00206] The greatest percentage of misclassifications obtained using invasion cluster 1 involved false positives, *i.e.*,  $17/38 = 44\%$  of the non-invasive tumors were mis-classified as having an expression profile associated with the invasive phenotype. To improve the overall accuracy of the method, the sample set was re-structured so as to include data only from the twelve invasive tumors correctly classified using invasion cluster 1, and from the seventeen tumors mis-classified as false positives. (The false positives were considered to be non-invasive tumors (as, in fact they were) in carrying out the method steps to generate the second reference set, the concordance set, and the minimum segregation set.) Using this set of twenty-nine samples, another second reference set was generated by using the ~~Affymetrix~~ AFFYMETRIX® MICRODB (version 3.0) and ~~Affymetrix~~ AFFYMETRIX® DATA MINING TOOLS (DMT) (version 3.0) data analysis software to identify genes that were differentially regulated in invasion group compared to non-invasive group of patients at the statistically significant level ( $p < 0.05$ ; Student T-test). Candidate genes were included in the second reference set if they were identified by the DMT software as having p values of 0.05 or less both for up-regulated and down-regulated genes. 458 genes were identified as being members of the second reference set

Please amend paragraph 217 as follows:

[00217] The methods of the invention were used along with the data reported by Singh, *et al.* (2002) to identify gene clusters capable of distinguishing tumor samples having a Gleason score of 6 or 7 (low grade tumors) from those having a Gleason score of 8 or 9 (high grade tumors). The same first reference set described above in part A was used to generate concordance and minimum segregation sets for Gleason score stratification. The second reference set was obtained following the procedures described above in part B, using the supplemental data reported in Singh, *et al.* (2002) for 46 low grade tumors and six high-grade tumors. Thus, the second reference set was generated by using the ~~Affymetrix~~ AFFYMETRIX® MICRODB

(version 3.0) and ~~Affymetrix~~AFFYMETRIX® DATA MINING TOOLS (DMT) (version 3.0) data analysis software to identify genes that were differentially regulated in high grade group compared to low grade group of patients at the statistically significant level ( $p < 0.05$ ; Student T-test). Candidate genes were included in the second reference set if they were identified by the DMT software as having p values of 0.05 or less both for up-regulated and down-regulated genes. 2144 genes were identified as being members of the second reference set

Please amend paragraph 263 as follows:

[00263] Here we identified metastasis-associated gene expression signatures based on expression profiling human prostate carcinoma xenografts derived from the same highly metastatic variant implanted at orthotopic (metastasis promoting setting) and ectopic (metastasis suppressing setting) sites, demonstrating that distinct malignant behavior of highly metastatic cells associated with the site of inoculation in a nude mouse is dependent upon differential gene expression in prostate cancer cells implanted either orthotopically or ectopically. We utilized the ~~Affymetrix~~AFFYMETRIX® GENECHIP system to compare the expression profiles of 12,625 transcripts in highly metastatic variant PC-3MLN4 implanted at orthotopic (metastasis promoting setting) (“PC3MLN4OR”) and ectopic (metastasis suppressing setting) (“PC3MLN4SC”) sites. PC-3MLN4 tumors growing in orthotopic metastasis-promoting setting appear to dramatically over-express a set of genes with well-established invasion-activation functions (Figure 46). Changes in expression for each transcript are plotted as Log10Fold Change Average expression level in PC-3MLN4OR versus Average expression level in less metastatic parental PC3OR and PC3MOR (recurrence signatures) (Fig. 47A) or versus Average expression level in PC3PC-3MLN4SC (invasion signatures) (Fig. 47B) and Log10Fold Change Average expression level in aggressive (recurrent or invasive) versus Average expression level in corresponding non-aggressive (non-recurrent or non-invasive) clinical phenotypes. Expression

profiling of the 12,625 transcripts in the orthotopic and s.c. xenografts derived from the cell variants of the PC-3 lineage was carried out. Transcripts differentially expressed at the statistically significant level ( $p < 0.05$ ; T-test) in the orthotopic PC-3M-LN4 tumors compared to the s.c. tumors of the same lineage as well as orthotopic tumors derived from the less metastatic parental PC-3M and PC-3 cell lines were identified using the Affymetrix AFFYMETRIX® MICRODB and Affymetrix AFFYMETRIX® DMT software. Similarly, transcripts differentially regulated in the 8 recurrent versus 13 non-recurrent (Fig. 47A) or 26 invasive versus 26 non-invasive (Fig. 47B) human prostate tumors at the statistically significant level ( $p < 0.05$ ; T-test) were identified. The small clusters of genes exhibiting highly concordant gene expression patterns in the xenograft model and clinical setting were identified using the methods of the invention. ...

Please amend paragraph 275 as follows:

[00275] The methods of the invention were used along with the data reported by Singh, *et al.* (2002) to identify gene clusters associated with an invasive phenotype. These data were the supplemental data reported in Singh, D., Febbo, P.G., *et al.*, "Gene Expression Correlates of Clinical Prostate Cancer Behavior," *Cancer Cell* March 2002 1:203-209, incorporated herein by reference. The clinical human prostate tumor samples were divided into two groups, invasive and non-invasive, as reported in Singh, *et al.* (2002). Invasive phenotype was assessed by determining the presence or absence of positive surgical margins ("PSM") and positive or negative capsular penetration ("PCP"). The reference set was obtained following the procedures described above in part B, using the supplemental data reported in Singh, *et al.* (2002) for 26 invasive (identified as having positive surgical margins and/or positive capsular penetration) and 26 non-invasive (identified as having no evidence of positive surgical margins and/or positive capsular penetration) human prostate tumors. Thus, the first reference set was obtained by using

the ~~Affymetrix~~AFFYMETRIX® MICRODB (version 3.0) and ~~Affymetrix~~AFFYMETRIX® DATA MINING TOOLS (DMT) (version 3.0) data analysis software to identify genes that were differentially regulated in invasive group compared to non-invasive group of patients at the statistically significant level ( $p < 0.05$ ; Student T-test). Candidate genes were included in the first reference set if they were identified by the DMT software as having p values of 0.05 or less both for up-regulated and down-regulated genes. 114 genes were identified as being members of the reference set (Table 47).

Please amend paragraph 309 as follows:

[00309] ~~Affymetrix~~AFFYMETRIX® **Arrays.** The protocol for mRNA quality control and gene expression analysis was that recommended by ~~Affymetrix~~AFFYMETRIX® (see the ~~Affymetrix~~AFFYMETRIX® website). In brief, approximately one microgram of mRNA was reverse transcribed with an oligo(dT) primer that has a T7 RNA polymerase promoter at the 5' end. Second strand synthesis was followed by cRNA production incorporating a biotinylated base. Hybridization to ~~Affymetrix~~AFFYMETRIX® U95Av2 arrays representing 12,625 transcripts overnight for 16 h was followed by washing and labeling using a fluorescently labeled antibody. The arrays were read and data processed using ~~Affymetrix~~AFFYMETRIX® equipment and software as reported previously (18, 19).

Please amend paragraph 310 as follows:

[00310] **Data Analysis.** Detailed protocols for data analysis and documentation of the sensitivity, reproducibility and other aspects of the quantitative statistical microarray analysis using ~~Affymetrix~~AFFYMETRIX® technology have been reported (18, 19). 40-50% of the surveyed genes were called present by the ~~Affymetrix~~AFFYMETRIX® MICROARRAY SUITE 5.0 software in these experiments. The concordance analysis of differential gene expression across the data sets was performed using ~~Affymetrix~~AFFYMETRIX® MICRODB v. 3.0 and

DMT v.3.0 software as described earlier (18, 19). We processed the microarray data using the ~~Affymetrix~~AFFYMETRIX® MICROARRAY SUITE v.5.0 software and performed statistical analysis of expression data set using the ~~Affymetrix~~AFFYMETRIX® MICRODB and ~~Affymetrix~~AFFYMETRIX® DMT software. This analysis identified a set of 218 genes (91 up-regulated and 127 down-regulated transcripts) differentially regulated in tumors from patients with recurrent versus non-recurrent prostate cancer at the statistically significant level ( $p < 0.05$ ) defined by both T-test and Mann-Whitney test (Table 69). The concordance analysis of differential gene expression across the clinical and experimental data sets was performed using ~~Affymetrix~~AFFYMETRIX® MICRODB v. 3.0 and DMT v.3.0 software as described earlier (19). The Pearson correlation coefficient for individual test samples and appropriate reference standard was determined using the Microsoft Excel software as described in the signature discovery protocol.

Please amend paragraph 314 as follows:

[00314] Step 1. Sets of differentially regulated transcripts were independently identified for each experimental conditions (*see* below) and clinical samples using the ~~Affymetrix~~AFFYMETRIX® microarray processing and statistical analysis software package as described in this example[[s]]'s Materials and Methods section.

Please amend paragraph 315 as follows:

[00315] Step 2. Sub-sets of transcripts exhibiting concordant expression changes in clinical and experimental samples were identified using the ~~Affymetrix~~AFFYMETRIX® MICRODB and DMT software. Sub-sets of transcripts were identified with concordant changes of transcript abundance behavior in recurrent versus non-recurrent clinical tumor samples (218 transcripts) and experimental conditions independently defined for each signature (Signature 1: PC-3MLN4 orthotopic versus s.c. xenografts; Signature 2: PC-3MLN4 versus PC-3M & PC-3 orthotopic



xenografts; Signature 3: PC-3/LNCap consensus class, Glinsky, G.V., Kronen-Herzig, A., Glinskii, A.B., Gebauer, G. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. *Molecular Carcinogenesis*, 37: 209-221, 2003). Thus, from a set of 218 transcripts three concordant sub-sets of transcripts were identified corresponding to each binary comparison of clinical and experimental samples.

Please amend paragraph 321 as follows:

**[00321] Identification of molecular signatures distinguishing sub-groups of prostate cancer patients with distinct clinical outcomes after therapy.** To identify the outcome predictor signatures, we utilized as a training data set the expression analysis of 12,625 transcripts in 21 prostate tumor samples obtained from prostate cancer patients with distinct clinical outcomes after therapy. Using biochemical evidence of relapse after therapy as a criterion of treatment failure, 21 patients were divided into two sub-groups, representing prostate cancer with recurrent (8 patients) and non-recurrent (13 patients) clinical behavior (14). We processed the original U95Av2 GENECHIP CEL files using the ~~Affymetrix~~ AFFYMETRIX® Microarray Suite 5.0 software and performed statistical analysis of expression data set using the ~~Affymetrix~~ AFFYMETRIX® MICRODB and ~~Affymetrix~~ AFFYMETRIX® DMT software. This analysis identified a set of 218 genes (91 up-regulated and 127 down-regulated transcripts) differentially regulated in tumors from patients with recurrent versus non-recurrent prostate cancer at the statistically significant level ( $p < 0.05$ ) defined by both T-test and Mann-Whitney test (Table 70